

What's HyperSpy?

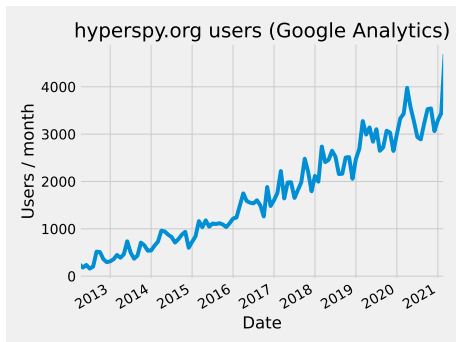
Francisco de la Peña



HyperSpy Workshop 2021
ePSIC Diamond Light Source (Cloud)
19th of April 2021

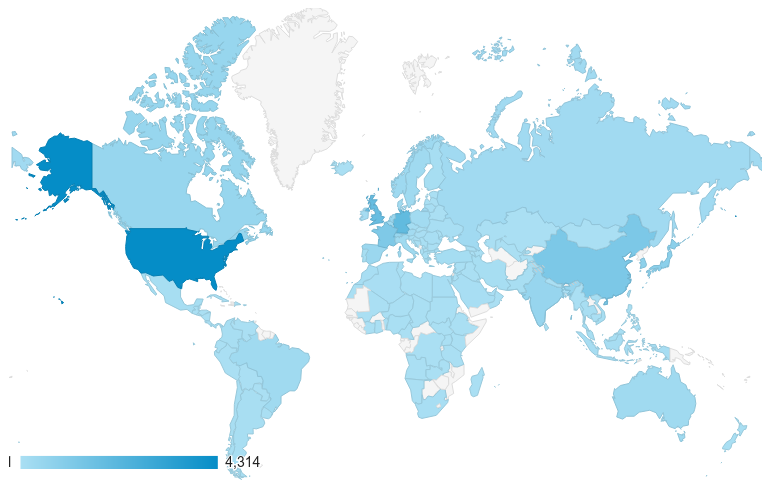
HyperSpy today—some stats

- > 3000 downloads/month?
- > 80k lines of code
- > 6000 citations
- > 50 contributors
- Used by more than 85 other packages in GitHub
- 300 GitHub Stars



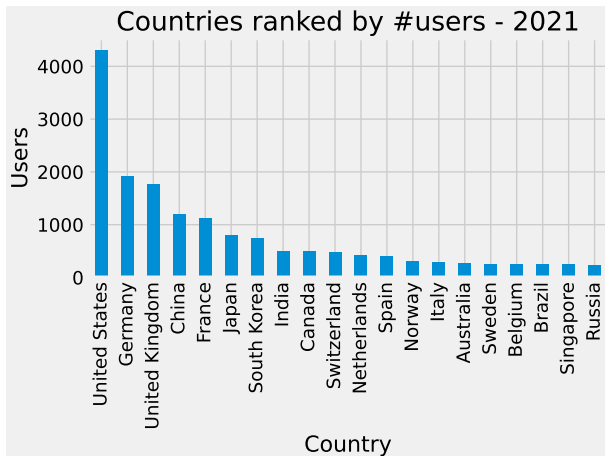
HyperSpy users by country

hyperspy.org users from April 2020-March 2021



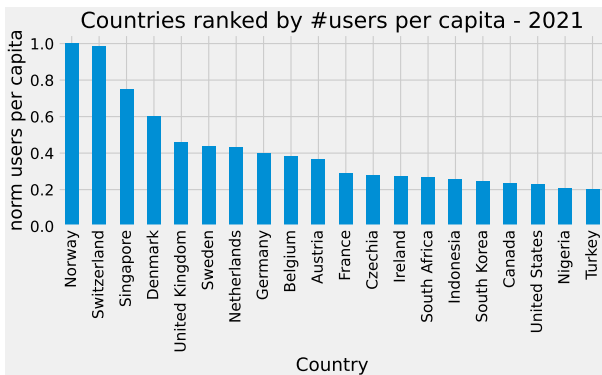
HyperSpy users by country

hyperspy.org users from April 2020-March 2021



HyperSpy users by country

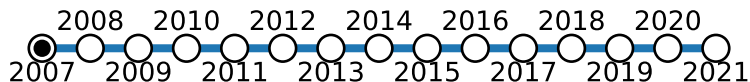
hyperspy.org users from April 2020-March 2021



What is HyperSpy? Take 1.

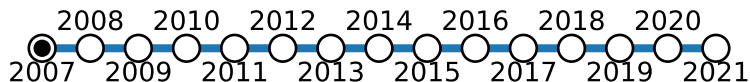
HyperSpy is a computer program for electron microscopy data analysis.

What is HyperSpy? (2007)



- During my PhD at the Paris-Sud University (LPS-Orsay), I started using Python for EELS data analysis.
- By placing the functions in classes a structure started to emerge.

What is HyperSpy? (2007)



- During my PhD at the Paris-Sud University (LPS-Orsay), I started using Python for EELS data analysis.
- By placing the functions in classes a structure started to emerge.

HyperSpy is ...

My personal set of Python routines for EELS data analysis of spectrum images.

What is HyperSpy? (2010)

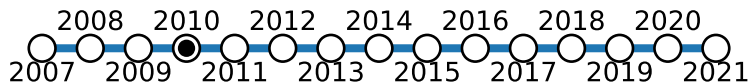


- Multivariate analysis routines implemented.
- EELSLab released under GPLv2 license.
- First users. First trainings in Paris and Oxford.

HyperSpy is ...

My personal set of Python routines for EELS data analysis of spectrum images.

What is HyperSpy? (2010)

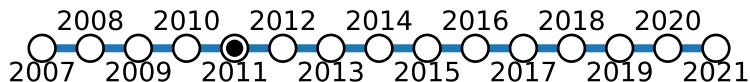


- Multivariate analysis routines implemented.
- EELSLab released under GPLv2 license.
- First users. First trainings in Paris and Oxford.

HyperSpy is ...

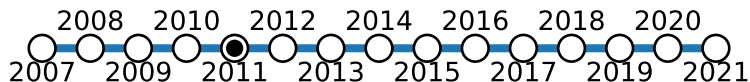
an *open-source* Python package for *data analysis of EELS multi-dimensional datasets*.

What is HyperSpy? (2011)



- Michael Sarahan (SuperSTEM) and Stefano Mazzucco (NIST) join the development team.
- We refactor the code to make it multi-dimensional.
- We rename it to “HyperSpy”.

What is HyperSpy? (2011)

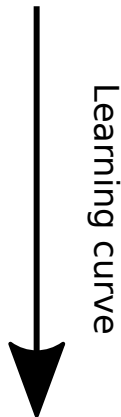
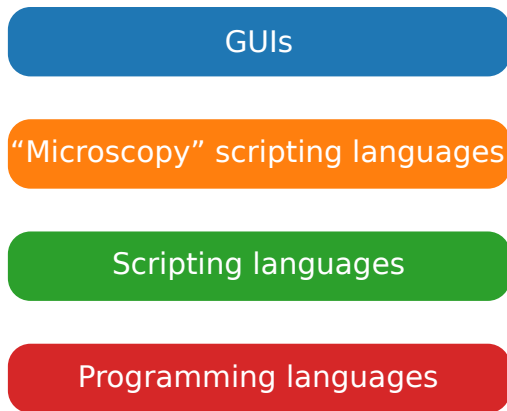


- Michael Sarahan (SuperSTEM) and Stefano Mazzucco (NIST) join the development team.
- We refactor the code to make it multi-dimensional.
- We rename it to “HyperSpy”.

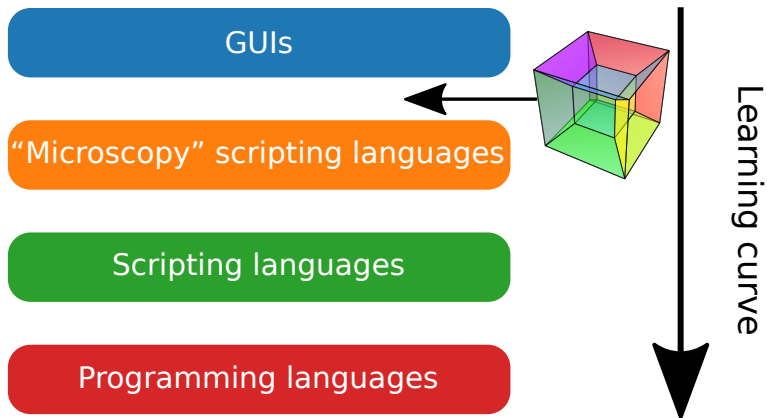
HyperSpy is ...

an open-source *open-development* Python package for *data analysis of multi-dimensional datasets*.

Why writing another software package?

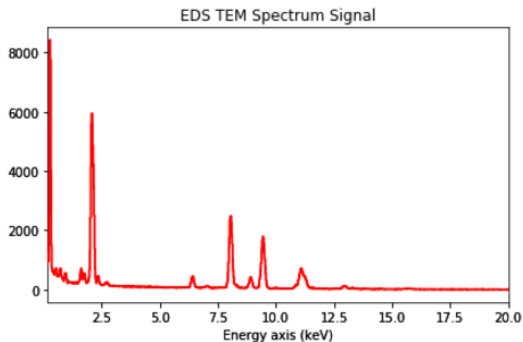


Why writing another software package?



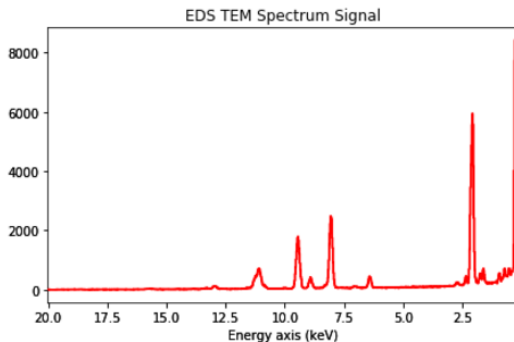
Example of HyperSpy vs Digital Micrograph syntax

```
In [4]: s.plot()
```



Example of HyperSpy vs Digital Micrograph syntax

```
In [5]: s.isig[::-1].plot()
```

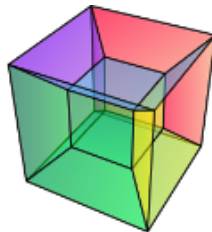


Example of HyperSpy vs Digital Micrograph syntax

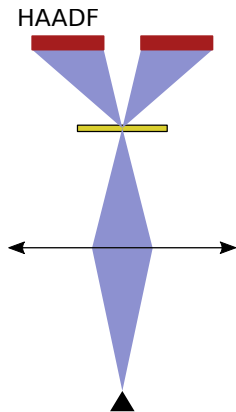
```
image img := GetFrontImage()
if (img.ImageGetNumDimensions() == 3)
{
    number sx = img.ImageGetDimensionSize(0)
    number sy = img.ImageGetDimensionSize(1)
    number sz = img.ImageGetDimensionSize(2)
    image res = Slice3(img, 0, 0, sz - 1, 0, sx, 1, 1, sy, 1, 2, sz, -1)
    res.ImageCopyCalibrationFrom(img)
    TagGroupCopyTagsFrom(ImageGetTagGroup(res), ImageGetTagGroup(img))
    res.ImageSetName(img.ImageGetName() + "_reversed")
    ShowImage(res)
}
```

HyperSpy design goals in 2011

- Cutting-edge features.
- Truly multi-dimensional.
- Scalable
 - Powerful, yet
 - easy to use.
 - easy to learn.
 - Easy to extend and contribute to.
 - Code fully available.

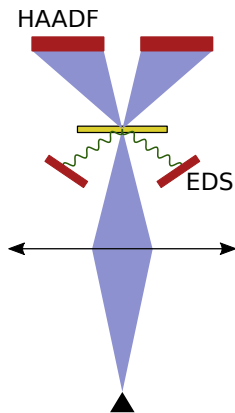


Why the emphasis on multi-dimensional datasets?



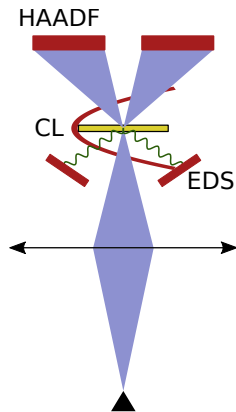
Detectors	Dimensions
HAADF	(x, y)

Why the emphasis on multi-dimensional datasets?



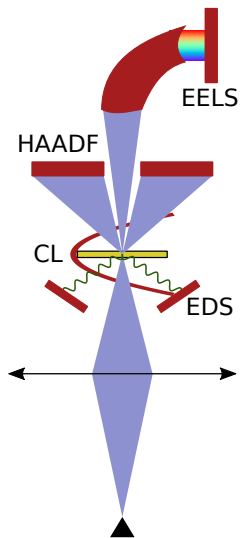
Detectors	Dimensions
HAADF	(x, y)
EDX	$(x, y \quad \quad E)$

Why the emphasis on multi-dimensional datasets?



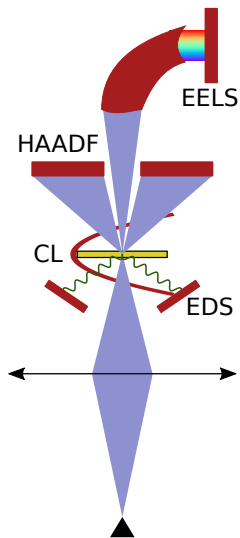
Detectors	Dimensions
HAADF	(x, y)
EDX	$(x, y \quad \quad E)$
CL	$(x, y \quad \quad \omega)$

Why the emphasis on multi-dimensional datasets?



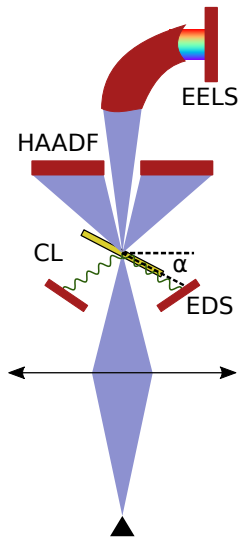
Detectors	Dimensions
HAADF	(x, y)
EDX	$(x, y \quad \quad E)$
CL	$(x, y \quad \quad \omega)$
EELS	$(x, y \quad \quad E)$

Why the emphasis on multi-dimensional datasets?



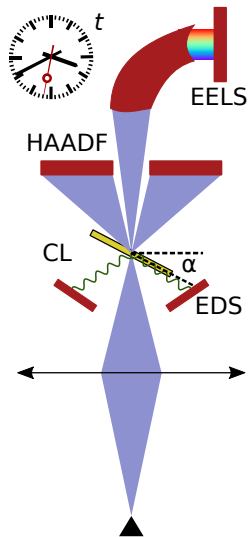
Detectors	Dimensions
HAADF	(x, y)
EDX	$(x, y \quad \quad E)$
CL	$(x, y \quad \quad \omega)$
EELS	$(x, y \quad \quad E)$
Diffraction	$(x, y \quad \quad x^*, y^*)$

Why the emphasis on multi-dimensional datasets?



Detectors	Dimensions
HAADF	(x, y)
EDX	$(x, y, \alpha E)$
CL	$(x, y \omega)$
EELS	$(x, y, \alpha E)$
Diffraction	$(x, y, \alpha x^*, y^*)$

Why the emphasis on multi-dimensional datasets?



Detectors	Dimensions
HAADF	(x, y)
EDX	$(x, y, \alpha, t E)$
CL	$(x, y, t \omega)$
EELS	$(x, y, \alpha, t E)$
Diffraction	$(x, y, \alpha, t x^*, y^*)$

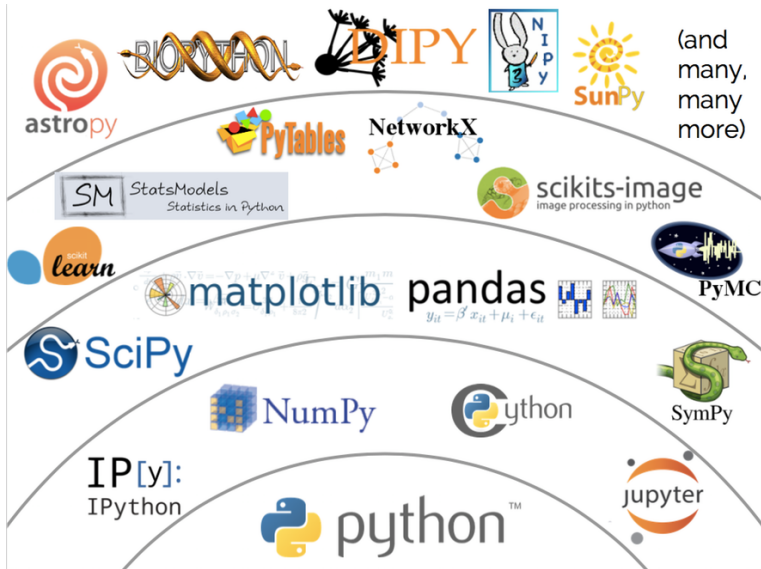
Why Python?

- Scripting language i.e. no need to compile \Rightarrow fast development.
- Genuine programming language.
- Readability: no need to know Python to understand what the code does \Rightarrow **Low entry barrier**.
- Runs natively on Windows, Mac OS, Linux.
- Open-source: not a black box and is free.

Why Python?

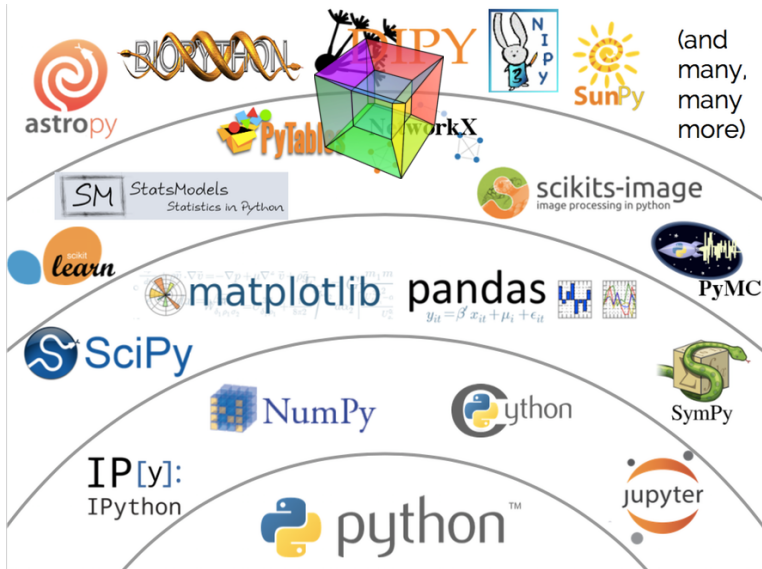
- Scripting language i.e. no need to compile \Rightarrow fast development.
- Genuine programming language.
- Readability: no need to know Python to understand what the code does \Rightarrow **Low entry barrier**.
- Runs natively on Windows, Mac OS, Linux.
- Open-source: not a black box and is free.
- **Lingua franca for scientific computing**.
- Unmatched environment for scientific computing:
 - Numpy+Scipy+matplotlib \geq (Matlab + Toolkits replacement)
 - Jupyter
 - sklearn, skimage...

The Scientific Python Stack



State of the "Scientific Python Stack" circa 2015, Jake VanderPlas

The Scientific Python Stack



State of the "Scientific Python Stack" circa 2015, Jake VanderPlas

Why Python?

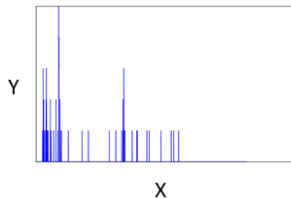


I came to Python lured by the language, but I stayed because of its [scientific] community.

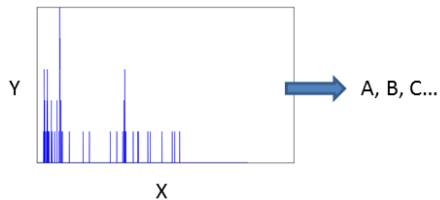
Fernando Pérez (UC Berkeley, creator of IPython)

Achieving sustainability: recycling

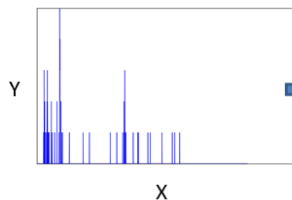
Achieving sustainability: recycling



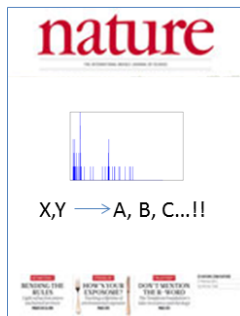
Achieving sustainability: recycling



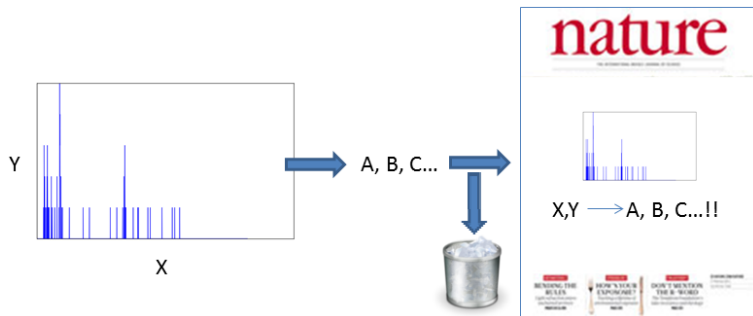
Achieving sustainability: recycling



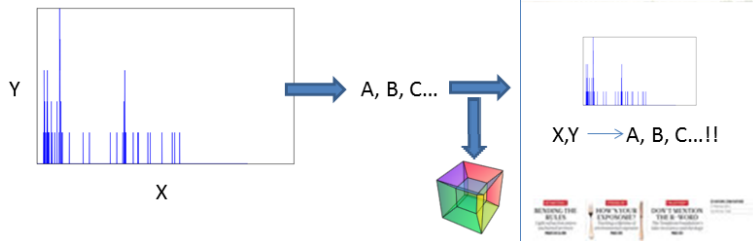
→ A, B, C... →



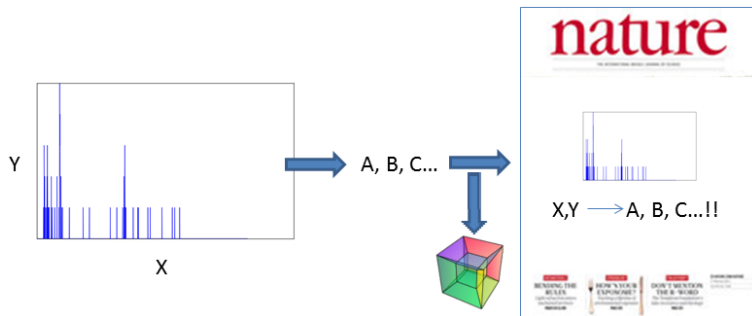
Achieving sustainability: recycling



Achieving sustainability: recycling



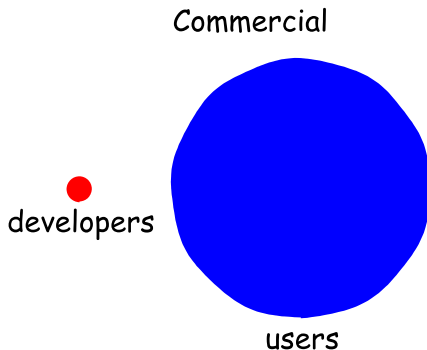
Achieving sustainability: recycling

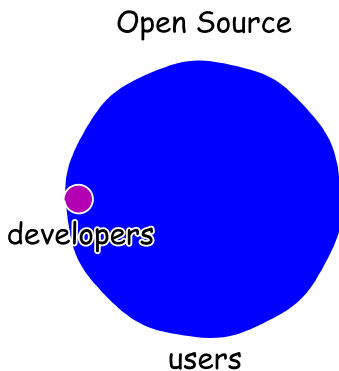


HyperSpy is ...

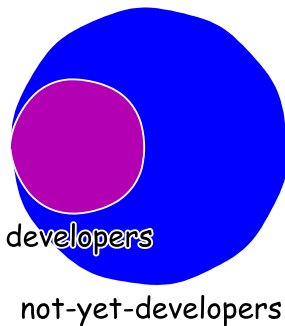
a *peer-reviewed open-access journal* specialized in code for EM multi-dimensional data analysis.

Sustainable opensource development





OpenSource Python





HyperSpy multi-dimensional data analysis

Fork me on GitHub

[Home](#) · [Download](#) · [Documentation](#) · [Demos](#) · [News](#) · [Support](#) · [Citing](#) · [Credits](#)

HyperSpy: multi-dimensional data analysis toolbox

HyperSpy is an open source Python library which provides tools to facilitate the interactive data analysis of multi-dimensional datasets that can be described as multi-dimensional arrays of a given signal (e.g. a 2D array of spectra a.k.a spectrum image).

HyperSpy aims at making it easy and natural to apply analytical procedures that operate on an individual signal to multi-dimensional arrays, as well as providing easy access to analytical tools that exploit the multi-dimensionality of the dataset.

Its modular structure makes it easy to add features to analyze different kinds of signals.

Highlights

- Two families of named and scaled axes: *signal* and *navigation*.
- Visualization tools for multi-dimensional spectra and images.

VERSIONS

Stable

1.3

[Download](#)

[Known issues](#)

Development

[Github](#)

SUPPORT

[Issue tracker](#)

[Mailing list](#)

OPEN CHAT



HyperSpy multi-dimensional data analysis

Fork me on GitHub

[Home](#) · [Download](#) · [Documentation](#) · [Demos](#) · [News](#) · [Support](#) · [Citing](#) · [Credits](#)

HyperSpy: multi-dimensional data analysis toolbox

HyperSpy is an open source Python library which provides tools to facilitate the interactive data analysis of multi-dimensional datasets that can be described as multi-dimensional arrays of a given signal (e.g. a 2D array of spectra a.k.a spectrum image).

HyperSpy aims at making it easy and natural to apply analytical procedures that operate on an individual signal to multi-dimensional arrays, as well as providing easy access to analytical tools that exploit the multi-dimensionality of the dataset.

Its modular structure makes it easy to add features to analyze different kinds of signals.

Highlights

- Two families of named and scaled axes: *signal* and *navigation*.
- Visualization tools for multi-dimensional spectra and images.

VERSIONS

Stable

1.3

[Download](#)

[Known issues](#)

Development

[Github](#)

SUPPORT

[Issue tracker](#)

[Mailing list](#)

OPEN CHAT

Community and support

This repository: `label:"release: next patch"` Pull requests Issues Marketplace Gist

hyperspy / hyperspy

Unwatch 18 Unstar 103 Fork 69

Code Issues 203 Pull requests 34 Projects 0 Wiki Settings Insights

Filters Labels Milestones [New issue](#)

Clear current search query, filters, and sorts

<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	92 Open	<input checked="" type="checkbox"/>	15 Closed	Author	Labels	Projects	Milestones	Assignee	Sort
<input type="checkbox"/>	<input checked="" type="checkbox"/>		Numpy Update Breaks MRC Reader Plugin	<code>release: next patch</code>	<code>type: bug</code>	#1705 opened 7 days ago by AndrewHerzing					2
<input type="checkbox"/>	<input checked="" type="checkbox"/>		Install dev version on windows	<code>release: next patch</code>	<code>type: bug</code>	#1704 opened 7 days ago by ericpre					
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Fix dm reader when microscope info missing	<code>release: next patch</code>	<code>status: needs review</code>	<code>type: bug</code>	<code>type: bug-fix</code>				2
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Fix a few bugs with rectangle and span roi and other tweaks.	<code>release: next patch</code>	<code>status: needs review</code>	<code>type: bug-fix</code>					5 of 5
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Events: Cannot clear connected functions	<code>affects: documentation</code>	<code>release: next patch</code>	<code>type: bug</code>					9
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Reverse plot_spectra legend order	<code>release: next patch</code>	<code>type: bug-fix</code>	#1695 by jat255 was merged 19 days ago					11
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Ordering of legend in plot_spectra	<code>release: next patch</code>	<code>status: fix-submitted</code>	<code>type: bug</code>					1
<input type="checkbox"/>	<input checked="" type="checkbox"/>		Rectangle roi not working with non-square pixel	<code>release: next patch</code>	<code>type: bug</code>	#1693 opened 25 days ago by ericpre					5

Community and support

This screenshot shows the GitHub interface for the repository 'hyperspy/hyperspy'. The top navigation bar includes 'This repository', 'label:"release: next patch"', 'Pull requests', 'Issues', 'Marketplace', and 'Gist'. The repository name 'hyperspy / hyperspy' is displayed, along with statistics: 'Unwatch 18', 'Unstar 103', and 'Fork 69'. Below this, there are tabs for 'Code', 'Issues 203', 'Pull requests 34', 'Projects 0', 'Wiki', 'Settings', and 'Insights'. A search bar contains 'is:pr is:open', and there are buttons for 'Filters', 'Labels', 'Milestones', and a green 'New pull request' button. The main content area shows a list of pull requests with columns for 'Author', 'Labels', 'Projects', 'Milestones', 'Reviews', 'Assignee', and 'Sort'. The list includes:

- #1702: Fixed small bug in holographic reconstruction during side-band search. X
- #1701: Make warnings about missing GUI features optional. ✓
- #1700: Fix dm reader when microscope info missing. X. Labels: release: next patch, status: needs review, type: bug. Type: bug-fix.
- #1699: Fix a few bugs with rectangle and span roi and other tweaks. ✓. Labels: release: next patch, status: needs review, type: bug-fix.
- #1698: ENH: Add "zero_fill" option to the background removal tool. X
- #1689: save(overwrite=False) now saves if no existing file. X. Labels: release: next patch, type: bug-fix.
- #1685: Fix Line2DROI. X. Labels: release: next patch, status: WIP.
- #1680: Model: custom optimised function and components. X. Labels: release: next minor, type: enhancement. Type: proposal.



HyperSpy

multi-dimensional data analysis

Fork me on GitHub

[Home](#) · [Download](#) · [Documentation](#) · [Demos](#) · [News](#) · [Support](#) · [Citing](#) · [Credits](#)

HyperSpy: multi-dimensional data analysis toolbox

HyperSpy is an open source Python library which provides tools to facilitate the interactive data analysis of multi-dimensional datasets that can be described as multi-dimensional arrays of a given signal (e.g. a 2D array of spectra a.k.a spectrum image).

HyperSpy aims at making it easy and natural to apply analytical procedures that operate on an individual signal to multi-dimensional arrays, as well as providing easy access to analytical tools that exploit the multi-dimensionality of the dataset.

Its modular structure makes it easy to add features to analyze different kinds of signals.

Highlights

- Two families of named and scaled axes: *signal* and *navigation*.
- Visualization tools for multi-dimensional spectra and images.

VERSIONS

Stable

1.3

[Download](#)

[Known issues](#)

Development

[Github](#)

SUPPORT

[Issue tracker](#)

[Mailing list](#)

OPEN CHAT



HyperSpy multi-dimensional data analysis

Fork me on GitHub

[Home](#) · [Download](#) · [Documentation](#) · [Demos](#) · [News](#) · [Support](#) · [Citing](#) · [Credits](#)

HyperSpy: multi-dimensional data analysis toolbox

HyperSpy is an open source Python library which provides tools to facilitate the interactive data analysis of multi-dimensional datasets that can be described as multi-dimensional arrays of a given signal (e.g. a 2D array of spectra a.k.a spectrum image).

HyperSpy aims at making it easy and natural to apply analytical procedures that operate on an individual signal to multi-dimensional arrays, as well as providing easy access to analytical tools that exploit the multi-dimensionality of the dataset.

Its modular structure makes it easy to add features to analyze different kinds of signals.

Highlights

- Two families of named and scaled axes: *signal* and *navigation*.
- Visualization tools for multi-dimensional spectra and images.

VERSIONS

Stable

1.3

[Download](#)

[Known issues](#)

Development

[Github](#)

SUPPORT

[Issue tracker](#)

[Mailing list](#)

OPEN CHAT

Community and support

hyperspy/hyperspy Multidimensional data analysis

Trond H @hentr Jul 21 06:23

hey, i am trying to remove specific frames from an image series, but cannot find a way to refer to the specific image in the image series, thinking something like this:

```
s.load('imageseries.dns')
for image in s:
    if np.median(image) < threshold:
        remove image from s

#or
for i in range(number_of_images)
    if s[image_number(i)]:
        remove s[image_number(i)] from s
```

but i cant figure out how to do something like `numpy.delete()` with `signal2D`

oh, sloppy should be

```
if np.median(s[image_number(i)]) < threshold:
```

on the second one

Francisco de la Peña @francisco-dlp Jul 21 06:27

No, you can't do that.

You could add the images that you want to keep to a list and then stack them together with `hs.stack`

A more efficient alternative would be to store the indices of the images that you want to keep in a list and then do the following:

```
s.data = s.data[idx_list, ...]
s.get_dimensions_from_data()
```

Michael Walls @mwalls Jul 21 07:09

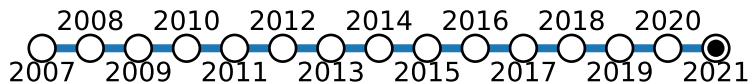
Hi Fran,

It's Laura, who is scared to go ion git, github etc. So I'm the intermediary. We tried what you suggested it gives this:

```
lauraboche$ pip install https://github.com/enthought/pyface/archive/master.zip
Invalid requirement: "
Traceback (most recent call last):
File "anaconda/lib/python3.6/site-packages/pip/_vendor/packaging/requirements.py", line 92, in init
req = REQUIREMENT_parseString(requirement_string)
File "anaconda/lib/python3.6/site-packages/pip/_vendor/pyarsing.py", line 1617, in parseString
raise exc
```

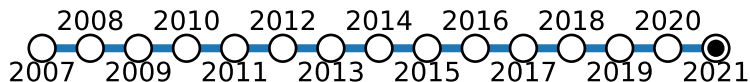
Click here to type a chat message. Supports GitHub flavoured markdown.

What is HyperSpy? (2021)



- Mature code base
- Used by a growing number of external packages e.g.: pyXem, atomap, ParticleSpy...

What is HyperSpy? (2021)

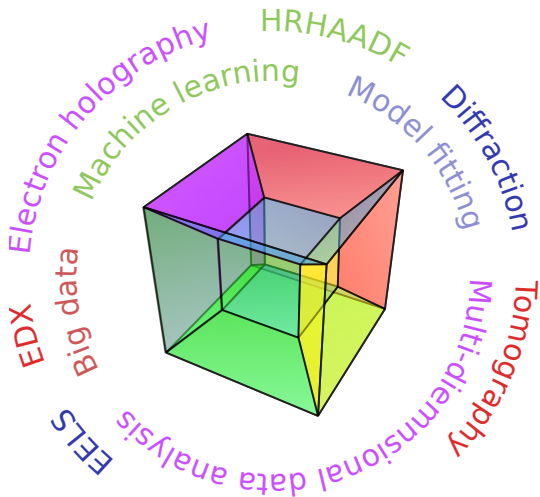


- Mature code base
- Used by a growing number of external packages e.g.: pyXem, atomap, ParticleSpy...

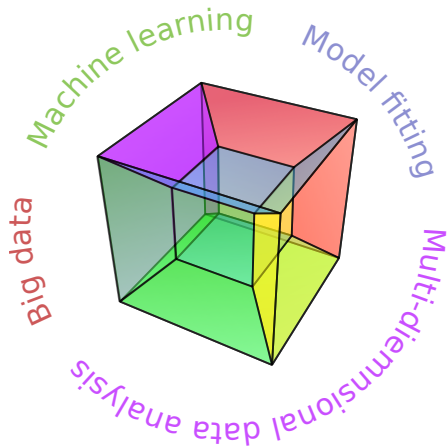
HyperSpy is ...

A *community* that aims at *pushing the boundaries* of data processing and analysis of *multi-dimensional datasets* across scientific fields by developing the features and *syntax* of the Python package of the same name.

Less is more: splitting HyperSpy



Less is more: splitting HyperSpy



Conclusion

When you people have a new idea:

Conclusion

When you people have a new idea:

- In academia: I hope no one scoops me

When you people have a new idea:

- In academia: I hope no one scoops me
- In Open Source: thanks goodness someone already thought of this!

Elizabeth Seiver @tweetotaler in Twitter

Thank you all for you attention

